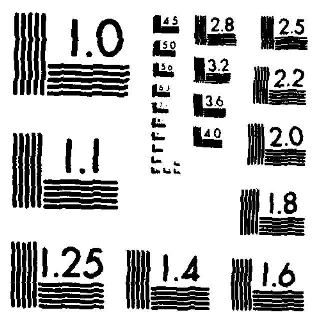
AD-A088 249 SOUTH CAROLINA UNIV COLUMBIA DEPT OF MATHEMATICS COM--ETC F/G 12/1
ASYMPTOTIC CONSERVATIENESS AND EFFICIENCY OF KRUSKAL-WALLIS TES--ETC(U)
MAY 80 L J WEI F49620-79-C-0140

UNCLASSIFIED AFOSR-TR-80-0566 NL

1 OF 1
AD
A088249

END
DATE
FILMED
9-80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AFOSR-TR- 80-0566

LEVEL II

AD A088249

ASYMPTOTIC CONSERVATIVENESS AND EFFICIENCY
OF KRUSKAL-WALLIS TEST FOR K DEPENDENT SAMPLES *

L. J. Wei

University of South Carolina

May, 1980

University of South Carolina
Department of Mathematics,
Computer Science, and Statistics
Columbia, South Carolina 29208

*Research supported by the United States Air Force Office of Scientific Research under Contract No. F49620-79-C-0140.

80 8 14 090

ASYMPTOTIC CONSERVATIVENESS AND EFFICIENCY

OF KRUSKAL-WALLIS TEST FOR K DEPENDENT SAMPLES

by L. J. Wei

Department of Mathematics, Computer Science, and Statistics

University of South Carolina - Columbia

ABSTRACT

The robustness (asymptotic conservativeness) of Kruskal-Wallis test under certain departures from mutual independence of K univariate samples is established. This robustness provides a procedure for testing the equality of K marginal distribution functions based on a broken random sample from a K-variate distribution which satisfies a mild condition. For the unbroken sample, a generalized Kruskal-Wallis test is proposed for testing the symmetry of a K-dimensional distribution function. The relative efficiency of the K-W test against the aligned rank order test is also examined under the normal shift model.

Key words and phrases: Broken sample; Asymptotically distribution free test; Positively dependent; Pitman efficiency.

AUTHOR'S FOOTNOTE

L. J. Wei is Associate Professor at the Department of Mathematics, Computer Science, and Statistics, University of South Carolina, Columbia, S.C. 29208. The author wishes to acknowledge helpful comments made by the referees and the Associate Editor.

1. INTRODUCTION

The problem of testing symmetry of a K-dimensional distribution function in its K arguments is discussed in many elementary nonparametric statistics books (c.f. Lehmann (1975), Hollander & Wolfe (1973)). One version of this problem is testing the hypothesis H_0 of no difference among K treatments in a randomized complete block design with random block effect or with fixed block effect based on aligned observations (c.f. Lehmann (1975), p. 270). However, if the sample is broken (c.f. DeGroot, Feder, and Goel (1971)), all the test procedures for testing H_0 are no longer valid. It is certainly interesting to see if anything can be done about this situation.

More specifically, suppose that a random sample $(Y_{11},\ldots,Y_{K1})',\ldots,(Y_{1n},\ldots,Y_{Kn})'$ of size n is drawn from a K-variate population with a continuous joint distribution function $G(y_1,\ldots,y_K)$ and marginal distribution functions F_i, $i = 1,\ldots,K$. The hypothesis which we wish to test is H_0: $F_1 \equiv F_2 \equiv \ldots \equiv F_K$. It is assumed that the distribution function G is symmetric in its K arguments under H_0. Now, however, suppose that before the sample values can be recorded, each observation vector $(Y_{1j},\ldots,Y_{Kj})'$ is broken into K separate components Y_{ij}, $i = 1,\ldots,K$ and the observations from each treatment i are available only in the form $Y_{ij_1},\ldots,Y_{ij_n}$, where $j_1,\ldots,j_n$ is some unknown permutation of $1,\ldots,n$. Since the blockings of $Y_{1j},\ldots,Y_{Kj}$ $(j = 1,\ldots,n)$ in the original sample are not known at the data analysis time, the observed values are called a broken random sample from the given K-variate population. This unfortunate situation is not uncommon in practice. An interesting example was cited by Hollander, Pledger, and Lin (1974).

The K-sample Kruskal-Wallis test (Kruskal (1952), Kruskal & Wallis (1952)) is not a distribution free test for the above situation; however, in this article we shall demonstrate that the K-W test is robust (asymptotic conservative) under a mild condition on the distribution G. That is, the asymptotic true level of significance of the K-W test for testing H_0 based on the broken sample is never greater than its preassigned nominal value, provided that the underlying joint distribution G is (pairwisely) positively quadrant dependent, i.e.,

$$P(Y_{ij} \le a,\ Y_{i'j} \le b) \ge P(Y_{ij} \le a)P(Y_{i'j} \le b),$$

for all $(a,b)' \in R^2$; $i,i' = 1,\ldots,K$; $i \ne i'$; $j = 1,\ldots,n$ (c.f. Lehmann (1966)). This has been shown to be a relatively weak form of positive dependence (c.f. Esary & Proschan (1972)). In practice, the positive dependence would be expected among observations within the same block in many situations. This is especially true for the randomized complete block design since subjects in the same block tend to be homogeneous. For K = 2, the K-W test is equivalent to the two-sample Wilcoxon test whose asymptotic conservativeness was studied by Hollander, Pledger and Lin (1974). Recently Rinott and Pollak (1978) have established the robustness of a class of test statistics which essentially are functionals of the difference between the empirical distribution functions of two positively dependent univariate samples.

In the second half of this article, a generalized Kruskal-Wallis test is proposed for testing H_0 based on the unbroken sample. This test allows all possible interblock comparisons and is shown to be asymptotically distribution free. Since the aligned rank order test also utilizes the information contained in the interblock comparisons and has been demonstrated to be more efficient than the Friedman test (c.f. Hodges & Lehmann (1962), Mehra & Sarangi (1967), and Sen (1968)), the asymptotic efficiency of the generalized K-W test with respect to the aligned rank order test is examined in this paper.

2. ASYMPTOTIC CONSERVATIVENESS OF KRUSKAL-WALLIS TEST

We define the scoring function ϕ for comparing two observations Y_{ij} and $Y_{i'j'}$ by

$$\phi(Y_{ij},Y_{i'j'}) = \begin{cases} 1, & Y_{ij} > Y_{i'j'} \\ -1, & Y_{ij} < Y_{i'j'} \\ 0, & \text{otherwise.} \end{cases}$$

Let $p_{ii'} = \int(2F_{i'}(y)-1)dF_i(y)$, where $i \neq i'$. Under H_0, $p_{ii'} = 0$. The ith component, W_{in}, of the vector score statistic $\underset{\sim}{W}_n$ is defined to be the total score comparing the ith sample with the ramining K-1 samples, i.e.,

$$W_{in} = \sum_{\substack{i'=1 \\ i'\neq i}}^{K} \sum_{j=1}^{n} \sum_{j'=1}^{n} (\phi(Y_{ij},Y_{i'j'})-p_{ii'})$$

and

$$\underset{\sim}{W}_n = (W_{1n},\ldots,W_{Kn})'.$$

<u>Theorem 1</u>. As $n \to \infty$, $n^{-3/2}\underset{\sim}{W}_n$ converges in distribution to a K-dimensional normal random vector with mean $\underset{\sim}{0}$ and covariance matrix $\Sigma = (\sigma_{ii'})$, where

$$\sigma_{ii'} = 4\,\mathrm{cov}\Big(\sum_{\substack{m=1 \\ m\neq i}}^{K} (F_m(Y_{i1})-F_i(Y_{m1})-p_{im}),\ \sum_{\substack{\ell=1 \\ \ell\neq i'}}^{K} (F_\ell(Y_{i'1})-F_{i'}(Y_{\ell 1})-p_{i'\ell})\Big),$$

$1 \le i,i' \le K$.

<u>Proof</u>. Let $U_{ii',n} = n^{-3/2} \sum_{j=1}^{n} \sum_{j'=1}^{n} (\phi(Y_{ij},Y_{i'j'})-p_{ii'})$,

$$\phi^0_{ii'}(y) = E\phi(y,Y_{i'j'})-p_{ii'} = 2F_{i'}(y) - 1 - p_{ii'}, \text{ and}$$

$$\phi^1_{ii'}(y) = E\phi(Y_{ij},y) - p_{ii'} = 1 - 2F_i(y) - p_{ii'}.$$

Also, let

$$g(a,b) = \phi(a,b) - p_{ii'} - \phi^0_{ii'}(a) - \phi^1_{ii'}(b) \text{ and}$$

$$U^*_{ii',n} = n^{-1/2}\left[\sum_{j=1}^{n} (\phi^0_{ii'}(Y_{ij}) + \phi^1_{ii'}(Y_{i'j}))\right]$$

$$= 2n^{-1/2} \sum_{j=1}^{n} (F_{i'}(Y_{ij}) - F_i(Y_{i'j}) - p_{ii'}).$$

Then

$$E(U_{ii',n} - U^*_{ii',n})^2 = n^{-3}E\left[\sum_{j=1}^{n}\sum_{j'=1}^{n} g(Y_{ij},Y_{i'j'})\right]^2$$

$$= n^{-3} \sum_{j=1}^{n}\sum_{j'=1}^{n}\sum_{k=1}^{n}\sum_{k'=1}^{n} h(j,j',k,k'), \qquad (2.1)$$

where $h(j,j',k,k') = E[g(Y_{ij},Y_{i'j'})g(Y_{ik},Y_{i'k'})]$. Because g is bounded, we can ignore any u terms of the sum in (2.1) if u is of order $o(n^3)$. Consider the following cases for which the number of terms is with order larger than or equal to $O(n^3)$ (where j,j',k, and k' represent four distinct indices):

(1) $h(j,j',k,k') = Eg(Y_{ij},Y_{i'j'})Eg(Y_{ik},Y_{i'k'}) = 0;$

(2) $h(j,j,k,k') = h(j,j',k,k) = 0;$

(3) $h(j,j',j,k') = h(j,j',k,j')$

$$= E\{E[g(Y_{ij},Y_{i'j'})g(Y_{ij},Y_{i'k'})|Y_{ij}]\}$$

$$= E\{E[g(Y_{ij},Y_{i'j'})|Y_{ij}]E[g(Y_{ij},Y_{i'k'})|Y_{ij}]\} = 0;$$

(4) $h(j,j',k,j) = h(j,j',j',k') = EE[g(Y_{ij},Y_{i'j'})g(Y_{ik},Y_{i'j})|Y_{ij},Y_{i'j}]$

$$= E\{E[g(Y_{ij},Y_{i'j'})|Y_{ij},Y_{i'j}]E[g(Y_{ik},Y_{i'j})|Y_{ij},Y_{i'j}]\} = 0.$$

It follows that $E(U_{ii',n} - U^*_{ii',n})^2 \to 0$, as $n \to \infty$. By Corollary 6 of Lehmann (1975, p. 289) and the fact that the vector $\langle U^*_{ii',n} \rangle$ has a $K(K-1)/2$-dimensional normal limiting distribution $N(\underset{\sim}{0},\Delta)$, where $\Delta = (\delta_{\ell\ell' kk'})$ and

$$\delta_{\ell\ell' kk'} = 4E(F_{\ell'}(Y_{\ell 1}) - F_{\ell}(Y_{\ell' 1}) - p_{\ell\ell'})(F_{k'}(Y_{k1}) - F_k(Y_{k'1}) - p_{kk'}),$$

$1 \le \ell,\ell',k,k' \le K$; $\ell \ne \ell'$; $k \ne k'$, the vector $\langle U_{ii',n} \rangle$ has the same multivariate normal limiting distribution $N(\underset{\sim}{0},\Delta)$. Therefore, $n^{-3/2}\underset{\sim}{W}_n$ has a multivariate normal limiting distribution with mean $\underset{\sim}{0}$ and covariance matrix Σ as is stated in the Theorem.

<u>Corollary 1</u>. If the joint distribution G of $(Y_{1j},\ldots,Y_{Kj})'$ is symmetric in its K arguments and F is the common marginal distribution function, then

$$\Sigma = \begin{pmatrix} \sigma^2 & \eta & \cdots & \eta \\ \eta & \sigma^2 & \cdots & \eta \\ \cdot & \cdot & \cdots & \cdot \\ \eta & \eta & \cdots & \sigma^2 \end{pmatrix}_{K \times K},$$

where $\sigma^2 = 4K(K-1)(1/12-\xi)$, $\eta = -4K(1/12-\xi)$, and $\xi = \operatorname{cov}(F(Y_{ij}), F(Y_{i'j}))$.

<u>Proof</u>. Let $V_{ij} = (K-1)F(Y_{ij}) - \sum_{\substack{\ell=1 \\ \ell \ne i}}^{K} F(Y_{\ell j})$. Then, it can be shown that

$\sigma^2 = 4\operatorname{var}(V_{ij}) = 4K(K-1)(1/12-\xi)$ and $\eta = 4\operatorname{cov}(V_{ij},V_{i'j}) = -4K(1/12-\xi)$.

The latent roots of Σ are $(\sigma^2 + (K-1)\eta) = 0$ with the latent vector $\underset{\sim}{q}_K = \frac{1}{\sqrt{K}}(1,1,\ldots,1)'$ and $(\sigma^2 - \eta)$ of multiplicity $(K-1)$ with $(K-1)$ orthonormal latent (any set of) vectors orthogonal to $\frac{1}{\sqrt{K}}(1,1,\ldots,1)'$, $\underset{\sim}{q}_1,\ldots,\underset{\sim}{q}_{K-1}$, say (c.f. Rao (1973), p. 67). In practice, the $\underset{\sim}{q}_i$ are perhaps most easily found by means of the Gram-Schmidt orthogonalization process. Since Σ is symmetric, therefore, $Q'\Sigma Q = \Lambda_{K\times K}$, where $Q = (\underset{\sim}{q}_1, \underset{\sim}{q}_2, \ldots, \underset{\sim}{q}_K)$, $\Lambda_{K\times K} = \mathrm{diag}(\lambda, \ldots, \lambda, 0)$ and $\lambda = \sigma^2 - \eta$. It follows that if G is symmetric in its K arguments,

$$S_n = n^{-3} \lambda^{-1} \sum_{i=1}^{K-1} [\underset{\sim}{q}_i' \underset{\sim}{W}_n]^2 \tag{2.2}$$

is asymptotically distributed as a chi-squared distribution with K-1 degrees of freedom. Note that if, for each j, $1 \le j \le n$, $Y_{1j}, \ldots, Y_{Kj}$ are mutually independent, then $\lambda = K^2/3$ and S_n in (2.2) becomes the usual K-sample Kruskal-Wallis test statistic. The next result gives the asymptotic conservativeness of the K-W test for testing H_0 under a certain dependent model.

<u>Corollary 2</u>. If G is symmetric in its K arguments and (pairwisely) positively quadrant dependent, then

$$\lim_{n\to\infty} P(3n^{-3} K^{-2} \sum_{i=1}^{K-1} [\underset{\sim}{q}_i' \underset{\sim}{W}_n]^2 \ge \chi^2_{1-\alpha}) \le \alpha, \text{ where } \chi^2_{1-\alpha}$$

is the $(1-\alpha)$th quantile of a chi-square random variable with K-1 degrees of freedom.

<u>Proof</u>. Since G is (pairwisely) positively quadrant dependent, it is easily seen that $\xi = \mathrm{cov}(F(Y_{ij}), F(Y_{i'j})) \ge 0$. Therefore, $\lambda = 4K^2(1/12-\xi) \le K^2/3$.

3. A GENERALIZED KRUSKAL-WALLIS TEST AND ITS EFFICIENCY

If the sample $(Y_{11}, \ldots, Y_{K1})', \ldots, (Y_{1n}, \ldots, Y_{Kn})'$ is not broken, the parameter λ in (2.2) can be consistently estimated from the data. One such consistent estimator $\hat{\lambda}$, say, for the case G is symmetric in its K arguments, is $K^2(1/3 - 4\hat{\xi})$, where

$$\hat{\xi} = \int(\hat{F}_n(x) - 1/2)(\hat{F}_n(y) - 1/2)d\hat{H}_n(x,y),$$

$$\hat{F}_n(x) = (nK)^{-1} \sum_{i=1}^{K} \sum_{j=1}^{n} I(Y_{ij} \leq x),$$

$$\hat{H}_n(x,y) = 2(nK(K-1))^{-1} \sum_{i \neq i'}\sum \sum_{j=1}^{n} I(Y_{ij} \leq x, Y_{i'j} \leq y),$$

and I(A) is the indicator function of the event A. The consistency of the estimator $\hat{\xi}$ can easily be obtained through the fact that $\sup_x|\hat{F}_n(x)-F(x)| \xrightarrow[n\to\infty]{a.s.} 0$, $\sup_{(x,y)} |\hat{H}_n(x,y) - H(x,y)| \xrightarrow[n\to\infty]{a.s.} 0$, and the continuity of F, where H(x,y) is the joint distribution function of Y_{ij} and $Y_{i'j}$, $1 \leq i \neq i' \leq K$, under the condition of symmetry of G. It follows that the test, $\hat{S}_n = n^{-3}\hat{\lambda}^{-1}\sum_{i=1}^{K-1}(q_i'W_n)^2$, which allows interblock comparisons, is an asymptotically distribution free test for testing the symmetry of the joint distribution function G.

It is certainly interesting to examine the relative efficiency of the gernalized K-W test based on $\hat{S}_n$ with respect to the powerful aligned rank order test. For simplicity, we only consider the case K = 2. In this situation, the aligned rank order test is asymptotically equivalent to the signed-rank Wilcoxon test (c.f. Lehmann (1975), p. 278). Under the normal shift model, i.e., $(Y_{1j},Y_{2j})'$ is a bivariate normal random vector with mean $(\mu+\theta,\mu)$ and covariance matrix $\begin{pmatrix}1,\rho\\ \rho,1\end{pmatrix}$, the Pitman efficiency of the signed-rank Wilcoxon test is $(3/(4\pi(1-\rho)))^{1/2}$ (c.f. Lehmann (1975), p. 374) for testing

$$H_0': \theta = 0 \quad \text{against} \quad H_1': \theta > 0. \tag{3.1}$$

When K = 2, the statistic $\hat{S}_n$ is equivalent to W_{1n}. By Theorem 1, $n^{-3/2}W_{1n}$ has a normal limiting distribution under H_0' and H_1' and satisfies four conditions of Lehmann ((1975), p. 371-372). For the above normal shift model, the Pitman efficiency of the test $\hat{S}_n$ for testing (3.1) can be shown to be $[3/(4\pi(1-12\xi))]^{1/2}$.

Since $\xi = -\frac{1}{2\pi}\arcsin(-\frac{\rho}{2})$, it follows that the relative efficiency of the K-W test $\hat{S}_n$ to the aligned rank test (or the signed-rank Wilcoson test) is $\frac{1-\rho}{1+\frac{6}{\pi}\arcsin(-\frac{\rho}{2})}$, which is greater than 1, if $-1 < \rho < 0$; less than 1, if $0 < \rho < 1$; equal to 1, if $\rho = -1, 0$; and undefined, if $\rho = 1$.

REFERENCES

(1) DeGroot, Morris H., Feder, Paul I. and Goel, Prem K. (1971), "Matchmaking", Ann. Math. Statist. 42, 578-593.

(2) Esary, J. D. and Proschan, F. (1972), "Relationships among some concepts of bivariate dependence" Ann. Math. Statist. 43, 651-655.

(3) Hodges, J. L., Jr. and Lehmann, E. L. (1962), "Rank methods for combination of independent experiments in the analysis of variance", Ann. Math. Statist. 33, 482-97.

(4) Hollander, Myles, Pledger, Gordon, and Lin, Pi-Ern (1974), "Robustness of the Wilcoxson test to a certain dependence between samples", Ann. Statist. 2, 177-181.

(5) Hollander, Myles, and Wolfe, Douglas A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons.

(6) Kruskal, William H. (1952), "A nonparametric test for the several sample problem", Ann. Math. Statist. 23, 525-540.

(7) __________, and Wallis, W. Allen (1952), "Use of ranks in one-criterion variance analysis", J. of Amer. Statist. Assoc. 47, 583-621.

(8) Lehmann, E. L. (1966), "Some concepts of dependence", Ann. Math. Statist. 37, 1137-1153.

(9) Lehmann, E. L. (1975), Nonparametrics: Statistical Methods Based on Ranks, San Francisco: Holden-day, Inc.

(10) Mehra, K. L. and Sarangi, J. (1967), "Asymptotic efficiency of certain rank tests for comparative experiments", Ann. Math. Statist. 38, 90-107.

(11) Rao, C. R. (1973), Linear Statistical Inference and Its Application, (2nd edition), New York: John Wiley & Sons.

(12) Rinott, Yosef and Pollak, Moshe (1978), "Robustness of two-sample test sizes to positive dependence between samples", Tech. Report, The Hebrew Univ. of Jerusalem.

(13) Sen, P. K. (1968), "On a class of aligned rank order tests in two-way layouts", Ann. Math. Statist. 29, 1115-1124.

18 AFOSR 19 TR-80-0566

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD A088249	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ASYPTOTIC CONSERVATIENESS AND EFFICIENCY OF KRUSKAL-WALLIS TEST FOR K DEPENDENT SAMPLES		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) L. J. Wei		8. CONTRACT OR GRANT NUMBER(s) 15 F49620-79-C-0140
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of South Carolina, Department of Mathematics, Computer Science, and Statistics		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 16 61102F 2304 17 A5 A1
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scienctific Research/NM Bolling AFB, Washington, D. C. 20332		12. REPORT DATE 11 May 1980
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 10 13		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Broken sample; Asymptotically distribution free test; Positively dependent; Pitman efficiency.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The robustness (asymptotic conservativeness) of Kruskal-Wallis test under certain departures from mutual independence of K univariate samples is established. This robustness provides a procedure for testing the equlity of K marignal distribution functions on a broken random sample from a K-variate distribution which satisfies a mild condition. For the unbroken sample, a generalized Kruskal-Wallis test is proposed for testing the symmetry of a K-dimensional distribution function. The relative efficiency of the K-W test against the aligned rank order test is also examined under the normal shift

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE 410442

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. continued:

model.

END

DATE
FILMED

9-80

DTIC